Assignment W45: OpenRefine

1. **Create a \*tidy\* spreadsheet/table listing the names of Danish monarchs with their birth- and death-date and start and end of their reign. They should be sortable by year of birth. Suitable source website is for example**[**here**](https://kongehuset.dk/monarkiet-i-danmark/kongerakken)**, but you can also use another source, provided you reference it. (Collaboration is welcome. Remember to attach this spreadsheet to Brightspace submission)**

This spreadsheet has been attached to Brightspace submission. I used another source on the Danish monarchs, which is referenced at the end of the assignment.

1. **Does OpenRefine alter the raw data during sorting and filtering?**

No, OpenRefine does not alter the raw data during sorting and filtering, which is one of the advantages of OpenRefine. Instead, it is possible to press the "Undo/Redo"-button, which will show you the changes you have done to the raw data. Here you will be able to return to your raw data, and then also return back to the changes you have made.

1. **Fix the**[**interviews dataset**](https://ndownloader.figshare.com/files/11502815)**in OpenRefine enough to answer this question: "Which two months are reported as the most water-deprived/driest by the interviewed farmer households?"**

Open OpenRefine and the project you're working on. Find the column called months\_no\_water. Press the arrow in the top of the column and choose Edit Cells --> Transform...

Insert the following regular expressions to remove left square brackets:

value.replace("[", "")

Click OK

Insert the following regex to remove single quote marks:

value.replace("'", "")

Click OK

Insert the following regex to remove right square brackets:

value.replace("]", "")

Insert the following regex to separate months with semicolons and removing spaces:

value.replace(" ", "")

Now we have removed superfluous characters from the column.

Now click the arrow in the months\_no\_water column, press Facet --> custom text facet...

Insert the following regex:

value.split(";")

Click OK

Now you will see the text facet box to the left. Click Sort by: count

This will show you that the two months reported as the most water-deprived/driest by the interviewed farmer households were October and September.

References:

Danmarkshistorien.dk, “Kongerækken, ca. 950-”, https://danmarkshistorien.dk/vis/materiale/kongeraekken